

# PATENT ABSTRACTS OF JAPAN

(11)Publication number : 10-011543

(43)Date of publication of application : 18.01.1998

(51)Int.Cl.

G06K 9/68  
G06K 9/62

(21)Application number : 08-167364

(71)Applicant : MATSUSHITA ELECTRIC IND CO LTD

(22)Date of filing : 27.08.1998

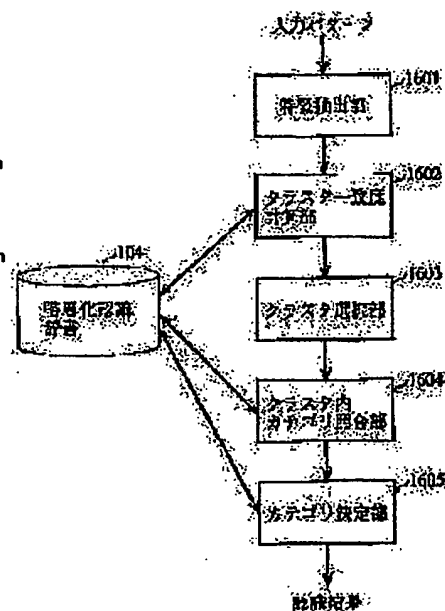
(72)Inventor : NAKAO ICHIRO  
TAKAKURA MINORU  
TAKENOUCHI MARIKO

## (54) PATTERN RECOGNITION DICTIONARY PRODUCTION DEVICE AND PATTERN RECOGNIZER

### (57)Abstract:

**PROBLEM TO BE SOLVED:** To shorten the time needed for the collation processing despite a large number of categories.

**SOLUTION:** A hierarchy recognition dictionary 104 registers a cluster standard feature vector representing a category sorted into clusters together with correspondence secured between the said category and its standard feature vector. A feature extraction part 1601 extracts a feature vector from an input pattern, and a cluster coincidence degree calculation part 1602 calculates a coincidence degree between the feature vector and the cluster standard feature vector. A cluster selection part 1603 selects a candidate cluster of a high coincidence degree, and an intra-cluster category collation part 1604 calculates a coincidence degree between the standard feature vector included in the candidate cluster and the feature vector. Then a category decision part 1605 decides a category of a high coincidence degree as a result of recognition.



### LEGAL STATUS

[Date of request for examination]

[Date of sending the examiner's decision of rejection]

[Kind of final disposal of application other than the examiner's decision of rejection or application converted registration]

[Date of final disposal for application]

[Patent number]

[Date of registration]

[Number of appeal against examiner's decision of rejection]

[Date of requesting appeal against examiner's decision of rejection]

[Date of extinction of right]

(19) 日本国特許庁 (JP)

(12) 公開特許公報 (A)

(11) 特許出願公開番号

特開平10-11543

(43) 公開日 平成10年(1998) 1月16日

| (51) Int. Cl. <sup>6</sup> | 識別記号  | 庁内整理番号  | FI           | 技術表示箇所  |
|----------------------------|-------|---------|--------------|---------|
| G 0 6 K 9/68               |       | 9061-5H | G 0 6 K 9/68 | E       |
| 9/62                       | 6 2 0 | 9061-5H | 9/62         | 6 2 0 A |

審査請求 未請求 請求項の数15 OL (全 19 頁)

(21) 出願番号 特願平8-167364

(22) 出願日 平成8年(1996) 6月27日

(71) 出願人 000005821

松下電器産業株式会社

大阪府門真市大字門真1006番地

(72) 発明者 中尾 一郎

大阪府門真市大字門真1006番地 松下電器  
産業株式会社内

(72) 発明者 高倉 穂

大阪府門真市大字門真1006番地 松下電器  
産業株式会社内

(72) 発明者 竹之内 磨理子

大阪府門真市大字門真1006番地 松下電器  
産業株式会社内

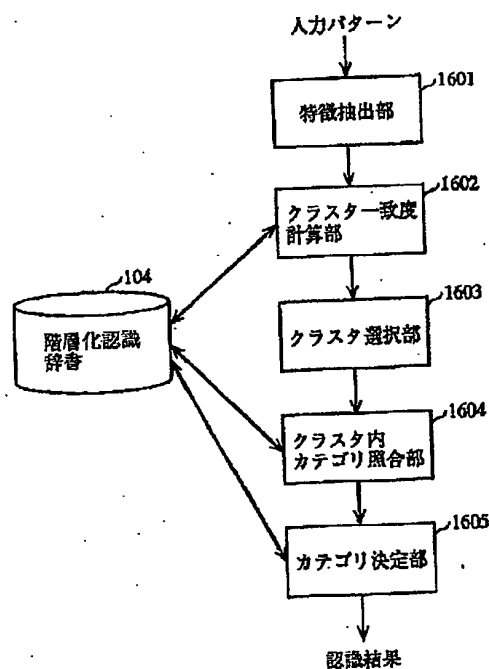
(74) 代理人 弁理士 中島 司朗

(54) 【発明の名称】 パターン認識用辞書作成装置及びパターン認識装置

(57) 【要約】

【課題】 従来のパターン認識装置においては、入力パターンの特徴ベクトルと全てのカテゴリとの標準特徴ベクトルとの照合処理を行うためカテゴリが多数の場合照合処理に長時間かかる。

【解決手段】 階層化認識辞書104は、クラスタに分類したカテゴリを代表するクラスタ標準特徴ベクトルと、クラスタに分類されたカテゴリとその標準特徴ベクトルとを対応付けて登録している。特徴抽出部1601は、入力パターンから特徴ベクトルを抽出し、クラスター一致度計算部1602は特徴ベクトルとクラスタ標準特徴ベクトルとの一致度を計算する。クラスタ選択部1603は一致度の高い候補クラスタを選択し、クラスタ内カテゴリ照合部1604は、候補クラスタ内のカテゴリの標準特徴ベクトルと特徴ベクトルとの一致度を計算する。カテゴリ決定部1605は一致度の高いカテゴリを認識結果として決定する。



## 【特許請求の範囲】

【請求項1】 カテゴリごとに入力される複数の学習用パターンからそのカテゴリの認識に用いる特徴量を抽出する特徴量抽出手段と、

前記特徴量抽出手段で抽出された特徴量に基づいて各カテゴリごとの標準特徴量を計算する標準特徴量計算手段と、

前記標準特徴量計算手段で計算された標準特徴量をクラスタ分析して、複数のクラスタに各カテゴリを分類するカテゴリ分類手段とを有するパターン認識用辞書作成装置であって、

前記カテゴリ分類手段で分類された各カテゴリの標準特徴量に基づいてクラスタのクラスタ標準特徴量を計算するクラスタ標準特徴量計算手段と、

前記クラスタ標準特徴量計算手段で計算されたクラスタのクラスタ標準特徴量と、前記カテゴリ分類手段で該クラスタに分類されたカテゴリと、前記標準特徴量計算手段で計算されたそのカテゴリの標準特徴量とを対応して階層化認識辞書に登録する登録手段とを備えることを特徴とするパターン認識用辞書作成装置。

【請求項2】 カテゴリごとに入力される複数の学習用パターンからそのカテゴリの認識に用いる特徴量を抽出する特徴量抽出手段と、

前記特徴量抽出手段で抽出された特徴量に基づいて各カテゴリごとの標準特徴量を計算する標準特徴量計算手段と、

前記標準特徴量計算手段で計算された標準特徴量をクラスタ分析して、複数のクラスタに分類する分類手段とを有するパターン認識用辞書作成装置であって、

前記分類手段で分類された標準特徴量に基づいて、各クラスタごとのクラスタ標準特徴量を計算するクラスタ標準特徴量計算手段と、

前記標準特徴量計算手段で計算されたカテゴリの標準特徴量と前記クラスタ標準特徴量計算手段で計算された各クラスタのクラスタ標準特徴量との一致度を計算する一致度計算手段と、

前記一致度計算手段で計算された一致度の値の最も高いクラスタにカテゴリを分類するカテゴリ分類手段と、

前記カテゴリ分類手段で分類されたカテゴリと該カテゴリの標準特徴量とを前記クラスタ標準特徴量とに対応して階層化認識辞書に登録する登録手段とを備えることを特徴とするパターン認識用辞書作成装置。

【請求項3】 前記階層化認識辞書に登録されているクラスタ内のカテゴリのうち一致度の値の最も低い値を該クラスタのクラスタ範囲と設定するクラスタ範囲設定手段と、

前記クラスタ範囲設定手段で設定されたクラスタのクラスタ範囲内に前記カテゴリ分類手段で他のクラスタに分類されたカテゴリがあるときには、階層化認識辞書の該クラスタに重複して該カテゴリとその標準特徴量とを対

応付けて追加登録する追加登録手段とを備えることを特徴とする請求項2記載のパターン認識用辞書作成装置。

【請求項4】 前記追加登録手段は、前記クラスタ標準特徴量計算手段で計算されたクラスタの標準特徴量と、前記標準特徴量計算手段で計算された各標準特徴量との一致度を計算する一致度計算部と、前記一致度計算部で計算された一致度が当該クラスタのクラスタ範囲内に前記カテゴリ分類手段で他のクラスタに分類された標準特徴量を有するカテゴリがあるか否かを判定するクラスタ範囲判定部と、

前記クラスタ範囲判定部があると判定したとき、当該カテゴリとその標準特徴量とを対応付けて当該クラスタに登録する重複登録部とを有することを特徴とする請求項3記載のパターン認識用辞書作成装置。

【請求項5】 カテゴリごとに入力される複数の学習用パターンからそのカテゴリの認識に用いる特徴量を抽出する特徴量抽出手段と、

前記特徴量抽出手段で抽出された特徴量に基づいて各カテゴリごとの標準特徴量を計算する標準特徴量計算手段と、

前記標準特徴量計算手段で計算された標準特徴量をクラスタ分析して、複数のクラスタに分類する分類手段とを有するパターン認識用辞書作成装置であって、

前記分類手段で分類された標準特徴量に基づいて、各クラスタごとのクラスタ標準特徴量を計算するクラスタ標準特徴量計算手段と、

前記特徴量抽出手段で抽出された特徴量と前記クラスタ標準特徴量計算手段で計算された各クラスタのクラスタ標準特徴量との一致度を計算する学習パターンクラスタ一致度計算手段と、

前記学習パターンクラスタ一致度計算手段で計算された一致度の最も高いクラスタに当該カテゴリと特徴量とを分類する特徴量分類手段と、

前記特徴量分類手段で分類された同一カテゴリの特徴量に基づいて各カテゴリごとの当該クラスタにおけるクラスタ内標準特徴量を計算するクラスタ内標準特徴量計算手段と、

前記カテゴリ分類手段で分類されたカテゴリと前記クラスタ内標準特徴量計算手段で計算された該カテゴリのクラスタ内標準特徴量とを前記クラスタ標準特徴量とに対応付けて階層化認識辞書に登録する登録手段とを備えることを特徴とするパターン認識用辞書作成装置。

【請求項6】 カテゴリごとに入力される複数の学習用パターンからそのカテゴリの認識に用いる特徴量を抽出する特徴量抽出手段と、

前記特徴量抽出手段で抽出された特徴量に基づいて各カテゴリごとの標準特徴量を計算する標準特徴量計算手段と、

前記標準特徴量計算手段で計算された標準特徴量をクラスタ分析して、複数のクラスタに分類する分類手段とを

有するパターン認識用辞書作成装置であって、  
前記分類手段で分類された標準特徴量に基づいて、各クラスごとにクラス標準特徴量を計算するクラス標準特徴量計算手段と、  
前記特徴量抽出手段で抽出された特徴量と前記クラス標準特徴量計算手段で計算された各クラスのクラス標準特徴量との一致度を計算する学習パターン-クラスター一致度計算手段と、  
前記学習パターン-クラスター一致度計算手段で計算された一致度の最も高いクラスに当該カテゴリを分類する特徴量分類手段と、  
前記特徴量分類手段で分類されたカテゴリと、該カテゴリの前記標準特徴量計算手段で計算された標準特徴量とを前記クラス標準特徴量に対応付けて階層化認識辞書に登録する登録手段とを備えることを特徴とするパターン認識用辞書作成装置。

【請求項7】 前記クラス標準特徴量計算手段は、同一クラス内の全カテゴリの標準特徴量の各要素を算術平均してクラス標準特徴量の各要素とすることを特徴とする請求項1、2、3、4、5又は6記載のパターン認識用辞書作成装置。

【請求項8】 上記一致度の計算は、市街化距離、二乗距離又は類似度のいずれかにより計算されることを特徴とする請求項2、3、4、5、6又は7記載のパターン認識用辞書作成装置。

【請求項9】 入力パターンをカテゴリとして認識するパターン認識装置であって、  
各カテゴリの学習用パターンの特徴量に基づいて計算された標準特徴量とカテゴリとを、当該標準特徴量のクラス分析によって複数のクラスに分類して、分類したクラスと、クラスに属するカテゴリの標準特徴量に基づいて計算されたクラス標準特徴量とに対応付けて登録している階層化認識辞書と、  
入力パターンから認識に用いる特徴量を抽出する特徴量抽出手段と、

前記特徴量抽出手段で抽出された特徴量と前記階層化認識辞書に登録されているクラス標準特徴量との一致度を計算するクラス一致度計算手段と、

前記クラス一致度計算手段で計算された一致度に基づいて候補クラスを選択する候補クラス選択手段と、  
前記候補クラス選択手段で選択された候補クラスに対応付けられた標準特徴量と前記特徴量抽出手段で抽出された特徴量との一致度を計算するカテゴリ一致度計算手段と、

前記カテゴリ一致度計算手段で計算された一致度に基づいて認識結果であるカテゴリを決定するカテゴリ決定手段とを備えることを特徴とするパターン認識装置。

【請求項10】 前記候補クラス選択手段は、  
前記クラス一致度計算手段で計算された一致度に基づいて、一致度の高いクラスから順に所定個数のクラス

タを候補クラスとして選択する所定数候補クラス選択部を有することを特徴とする請求項9記載のパターン認識装置。

【請求項11】 前記候補クラス選択手段は、  
前記クラス一致度計算手段で計算された一致度に基づいて、一致度が予め定めたしきい値以上のクラスを候補クラスとして選択するしきい値候補クラス選択部を有することを特徴とする請求項9記載のパターン認識装置。

10 【請求項12】 前記候補クラス選択手段は、  
前記クラス一致度計算手段で計算された一致度に基づいて、一致度の最も高いクラスの一致度を基準値として基準値との差又は比が一定値以内の一致度を有するクラスを候補クラスとして選択する基準値候補クラス選択部を有することを特徴とする請求項9記載のパターン認識装置。

【請求項13】 前記候補クラス選択手段は、  
前記クラス一致度計算手段で計算された一致度の高いクラスから順にクラス内に登録されているカテゴリ数が予め定めたカテゴリ数に到達するまでのクラスを候補クラスとして選択するカテゴリ数候補クラス選択部を有することを特徴とする請求項9記載のパターン認識装置。

【請求項14】 前記カテゴリ決定手段は、  
前記クラス選択手段で選択された候補クラスの前記クラス一致度計算手段で計算された一致度と、前記カテゴリ一致度計算手段で計算された一致度とに基づいて総合一致度を計算する総合一致度計算部と、  
前記総合一致度計算部で計算された総合一致度に基づいてカテゴリを決定する総合カテゴリ決定部とを有することを特徴とする請求項9、10、11、12又は13記載のパターン認識装置。

【請求項15】 階層化認識辞書を用いて入力パターンをカテゴリとして認識するパターン認識方法が記録された記録媒体であって、  
各カテゴリの学習用パターンの特徴量に基づいて計算された標準特徴量とカテゴリとを、当該標準特徴量のクラス分析によって複数のクラスに分類して、分類したクラスと、クラスに属するカテゴリの標準特徴量に基づいて計算されたクラス標準特徴量とに対応付けて登録している階層化認識辞書と、  
以下の特定ステップを実行するプログラムとが記録され、

特定ステップには、

入力パターンから認識に用いる特徴量を抽出する特徴量抽出ステップと、

前記特徴量抽出ステップで抽出された特徴量と前記階層化認識辞書に登録されているクラス標準特徴量との一致度を計算するクラス一致度計算ステップと、

50 前記クラス一致度計算ステップで計算された一致度

基づいて候補クラスタを選択する候補クラスタ選択ステップと、  
前記候補クラスタ選択ステップで選択された候補クラスタに対応付けられた標準特徴量と前記特徴量抽出ステップで抽出された特徴量との一致度を計算するカテゴリ一致度計算ステップと、  
前記カテゴリ一致度計算ステップで計算された一致度に基づいて認識結果であるカテゴリを決定するカテゴリ決定ステップとを含むことを特徴とする記録媒体。

#### 【発明の詳細な説明】

#### 【0001】

【発明の属する技術分野】本発明は、文字・図形等のパターン認識に用いる辞書の作成装置及びパターン認識装置に関する。

#### 【0002】

【従来の技術】従来からのパターン認識装置は、入力されたパターンから特徴ベクトルと呼ばれるパターンを認識するための特徴量である複数の数値データを抽出する。次に抽出した特徴ベクトルと、予め用意された各カテゴリの標準特徴ベクトルとを照合し、その一致度（距離、類似度等）を求める。ここで、カテゴリとは、例えば、文字認識の場合には識別字種をいい、認識対象の入力パターンの認識結果となる区分種類をいう。一致度の最も高い標準特徴ベクトルに対応するカテゴリを入力パターンの認識結果としている。このため、各カテゴリの標準特徴ベクトルを用意するため、各カテゴリごとに多数の学習用パターンから特徴ベクトルを抽出し、各カテゴリごとにその特徴ベクトルの平均値等を求めている。

#### 【0003】

【発明が解決しようとする課題】ところが、上記のようなパターン認識装置では、入力パターンから抽出された特徴ベクトルと認識結果となる全てのカテゴリの標準特徴ベクトルとを逐一照合しなければならないため、照合処理に長時間を要する不都合がある。特に、日本語文字認識では、漢字を含む多くのカテゴリがあるため、その照合処理に要する時間の短縮がパターン認識装置の実用上の課題となっている。

【0004】本発明は、上記課題に鑑み、照合処理に要する時間を大幅に短縮できるパターン認識装置及びその装置に用いるパターン認識辞書を作成する装置を提供することを目的とする。

#### 【0005】

【課題を解決するための手段】上記課題を解決するため、本発明は、カテゴリごとに入力される複数の学習用パターンからそのカテゴリの認識に用いる特徴量を抽出する特徴量抽出手段と、前記特徴量抽出手段で抽出された特徴量に基づいて各カテゴリごとの標準特徴量を計算する標準特徴量計算手段と、前記標準特徴量計算手段で計算された標準特徴量をクラスタ分析して、複数のクラスタに各カテゴリを分類するカテゴリ分類手段とを有す

るパターン認識用辞書作成装置であって、前記カテゴリ分類手段で分類された各カテゴリの標準特徴量に基づいてクラスタのクラスタ標準特徴量を計算するクラスタ標準特徴量計算手段と、前記クラスタ標準特徴量計算手段で計算されたクラスタのクラスタ標準特徴量と、前記カテゴリ分類手段で該クラスタに分類されたカテゴリと、前記標準特徴量計算手段で計算されたそのカテゴリの標準特徴量とを対応して階層化認識辞書に登録する登録手段とを備えることとしている。これによって、階層化認識辞書には、クラスタごとのクラスタ標準特徴量と、カテゴリと該カテゴリの標準特徴量とを対応して登録するので、この階層化認識辞書を用いたパターン認識装置では、先ず認識対象の入力パターンから抽出したパターン特徴量とクラスタ標準特徴量とを照合し、その一致度の高いクラスタに含まれるカテゴリの標準特徴量とパターン特徴量とを照合するだけで、全てのカテゴリの標準特徴量との照合をする必要がなくなるので、照合時間を短縮して、認識結果となるカテゴリを選択することができる。

#### 【0006】

【発明の実施の形態】以下、本発明に係るパターン認識用辞書作成装置及びパターン認識装置の実施の形態を図面を用いて説明する。

（実施の形態1）図1は、本発明に係るパターン認識用辞書作成装置の実施の形態1の構成図である。

【0007】パターン認識用辞書作成装置は、特徴抽出部101と、標準特徴計算部102と、クラスタ代表辞書作成部103と、階層化認識辞書104と、クラスタ内カテゴリ設定部105と、クラスタ内辞書作成部106とを備えている。特徴抽出部101は、認識結果となる各カテゴリごとに複数の学習用パターン入力を受けると、各学習用パターンからカテゴリの認識に用いる特徴量である複数の数値データからなる特徴ベクトルを抽出し、カテゴリに対応して特徴ベクトルを記憶する。

【0008】図2は、文字を認識結果とする入力された学習用パターンの一例を示す図である。このように文字を認識結果とする場合には、カテゴリは、各文字に対応する文字コードで示される。文字「あ」の文字コード201に対応して複数の文字画像「あ」202、203、…、204が学習用パターンとして入力されている。なお、本実施の形態では、文字を認識結果としているけれども、他の図形等を認識結果とする場合でも同様である。

【0009】文字コード201に対応する学習用パターンである文字画像「あ」202、203、…、204の特徴ベクトルを抽出すると、図3に示ようになる。なお、特徴ベクトルの抽出については、特開昭63-34682号公報に記載されている。ここで、文字コード201に対応する学習用パターンがM個あるとすると、特徴ベクトルもF1、F2、…、FMのM個が抽出され

る。また特徴ベクトルがn個の数値データからなるときは、例えば、文字画像204から抽出された特徴ベクトルFMは、

$$FM = (fM-1, fM-2, \dots, fM-n)$$

となる。

【0010】認識結果となるカテゴリである文字コードの総数がNである場合には、各カテゴリごとの特徴ベクトルの一覧301、302、303、…の総数はNとなる。特徴抽出部101は、N個のカテゴリについて、特徴ベクトルを全て抽出すると、標準特徴計算部102を起動する。標準特徴計算部102には、特徴抽出部101に起動されると、特徴抽出部101に記憶されている図3に示したような、各カテゴリごとの特徴ベクトルを読み出し、各カテゴリの標準特徴ベクトルを計算する。例えば、図2に示した文字コード201の文字「あ」の標準特徴ベクトルD = (d1, d2, ..., dn)は、次式の(数1)によってその要素djが求められる。

【0011】

【数1】

$$d_j = \frac{1}{M} \sum_{i=1}^M f_{i-j} \quad (j=1, \dots, n)$$

同様に、認識結果となる全てのカテゴリについて標準特徴ベクトルを計算して記憶すると、クラスタ代表辞書作成部103を起動する。クラスタ代表辞書作成部103は、標準特徴計算部102に起動されると、標準特徴計算部102に記憶されているカテゴリの数Nと、N個のカテゴリの標準特徴ベクトルを読み出してクラスタ分析の手法（「多変量統計解析法」田中豊、脇本和昌著、現代数学社 等に記載されている。）によって、複数のクラスタに各標準特徴ベクトルを分類する。次に各クラスタに分類された標準特徴ベクトルに基づいてクラスタの代表となるクラスタ標準特徴ベクトルを計算する。このクラスタ標準特徴ベクトルは、上記標準特徴計算部102と同様に、クラスタ標準特徴ベクトルの各要素を次式の(数2)によって計算し、クラスタ標準特徴ベクトルC = (c1, c2, ..., cn)を階層化認識辞書104にそのクラスタ番号とともに書き込む。

【0012】

【数2】

$$C_j = \frac{1}{R} \sum_{i=1}^R d_{i-j} \quad (j=1, \dots, n)$$

ここで、このクラスタに含まれるカテゴリの数は、R個であるものとする。全て（例えばP個）のクラスタについて、クラスタ標準特徴ベクトルが求められると、クラスタ内カテゴリ設定部105を起動する。階層化認識辞書104は、RAM等からなり、クラスタ代表辞書作成部103によって、クラスタ番号とクラスタ標準特徴ベクトルとを書き込まれる。

【0013】更に、クラスタ内カテゴリ設定部105によって、各クラスタに分類される各カテゴリが、追加分類されるものを含めて書き込まれ、クラスタ内辞書作成部106によって、クラスタに分類されているカテゴリに対応付けてその標準特徴ベクトルが書き込まれる。クラスタ内カテゴリ設定部105は、図4にその詳細な構成図を示すように、クラスター一致度計算部401と、クラスタ内カテゴリ登録部402と、クラスタ範囲設定部403と、クラスタ拡張部404とを備えている。

10 【0014】クラスター一致度計算部401は、クラスタ代表辞書作成部103に起動されると、標準特徴計算部102に記憶されている各カテゴリの標準特徴ベクトルD = (d1, d2, ..., dn)を1つずつ読み出し、階層化認識辞書104から各クラスタ標準特徴ベクトルC = (c1, c2, ..., cm)を読み出して、両者の特徴ベクトルの一致度を式(数3)、(数4)、(数5)のいずれかを用いて計算する。

【0015】

【数3】

$$L1 = \sum_{i=1}^n |d_i - c_i|$$

【0016】

【数4】

$$L2 = \sum_{i=1}^n (d_i - c_i)^2$$

【0017】

【数5】

$$S = \left( \sum_{i=1}^n (d_i \times c_i) \right) / (|D| \times |C|)$$

$$\text{ここで } |D| = \left( \sum_{i=1}^n (d_i^2) \right)^{1/2}$$

$$|C| = \left( \sum_{i=1}^n (c_i^2) \right)^{1/2}$$

式(数3)は市街化距離L1を一致度を用いるものであり、L1の値の小さいもの程一致度が高くなる。式(数4)は、二乗距離L2を用いるものであり、L2の値の小さいもの程一致度が高くなる。式(数5)は、類似度Sを用いるものであり、Sの値の大きいもの程一致度が高くなる。本実施の形態では、式(数3)の市街化距離L1を用いて一致度を計算する。得られた一致度をクラスタ内カテゴリ登録部402に通知する。

40 【0018】クラスタ内カテゴリ登録部402は、クラスター一致度計算部401から通知された一致度の最も高い（市街化距離L1の最小の）クラスタに標準特徴ベクトルを有するカテゴリを分類する。即ち、標準特徴ベクトルで特徴付けられるカテゴリを例えばp個のクラスタのいずれかに分類して、階層化認識辞書104に登録するとともに、そのカテゴリと一致度とをクラスタに対応付けて記憶する。図5は、クラスタ内カテゴリ登録部4

02に記憶されているクラスタ501と分類されたカテゴリ502とその一致度503とを示している。例えば、クラスタ1の場合には、文字コード「あ」、「お」、…、「め」のm1個のカテゴリが分類されて、各カテゴリのクラスタ1のクラスタ標準特徴との一致度H1-1, H1-2, …, H1-m1が記憶されている。なお、ここでは市街化距離L1として計算された値の最小値を一致度Hとしている。全てのカテゴリをクラスタに分類すると、クラスタ範囲設定部403を起動する。

【0019】クラスタ範囲設定部403は、クラスタ内カテゴリ登録部402に起動されると、クラスタ内カテゴリ登録部402に記憶されている各クラスタに分類されたカテゴリの一致度の最も低いカテゴリとの一致度をクラスタ範囲として設定する。上述したクラスタ1の場合には、クラスタ範囲r1は、H1-1, H1-2, …, H1-m1の最大値(市街化距離であるので最大値が一致度の最も低いものとなる。)

$r1 = \max(H1-1, H1-2, \dots, H1-m1)$ に設定される。同様にクラスタ2, …, クラスタpについてもクラスタ範囲r2, …, rpが設定される。

【0020】全てのクラスタについてクラスタ範囲rを設定すると、クラスタ拡張部404に各クラスタのクラスタ範囲rを通知する。クラスタ拡張部404は、クラスタ範囲設定部403からクラスタ範囲rの通知を受けると、階層化認識辞書104からクラスタ標準特徴ベクトルCを読み出す。読み出した各クラスタ標準特徴ベクトルCと、標準特徴計算部102に記憶されている全てのカテゴリの標準特徴ベクトルDとの一致度を式(数3)を用いて計算し、 $L1 \leq r$ となるカテゴリを階層化認識辞書104のクラスタに分類して追加記憶させる。

【0021】即ち、図6に示すように、クラスタ内カテゴリ登録部402において、クラスタjに分類されていたカテゴリである文字コード「軍」は、クラスタiのクラスタ範囲ri内に存在するので、クラスタiにも重複して分類する。これによって、後述する図7に示すように、クラスタiとクラスタjのクラスタ内カテゴリ欄701に文字コード「軍」が重複して登録される。全て(P個)のクラスタについて、全て(N個)のカテゴリがクラスタ範囲ri内に存在するか否かの判断が終了すると、クラスタ内辞書作成部106を起動する。

【0022】クラスタ内辞書作成部106は、クラスタ拡張部404に起動されると、階層化認識辞書104に記憶されているクラスタに分類されたカテゴリを読み出し、各カテゴリの標準特徴ベクトルを標準特徴計算部102から読み出して、階層化認識辞書104にクラスタ内カテゴリに対応付けて記憶させる。図7は、このようにして階層化認識辞書104に記憶された記憶内容を示す説明図である。

【0023】この階層化認識辞書104は、クラスタ欄

702とクラスタ標準特徴ベクトル欄703とクラスタ内カテゴリ欄701と標準特徴ベクトル欄704とを含んでいる。クラスタ欄702にはクラスタ番号「1」～「p」を付されたクラスタが記憶され、クラスタ標準特徴ベクトル欄703にはそのクラスタを代表するクラスタ標準特徴ベクトルCが記憶されている。クラスタ標準特徴ベクトルCのサフィックスは、クラスタ番号を示し、クラスタ標準特徴ベクトルCの要素cの()内の第1文字はクラスタ番号を示し、第2文字は特徴要素の番号を示している。

【0024】クラスタ内カテゴリ欄701には、各クラスタごとにクラスタに分類されるカテゴリが対応付けて記憶されている。ここでカテゴリは、上述したように文字コードで示されている。標準特徴ベクトル欄704には、クラスタ内カテゴリ欄701のカテゴリと対応付けて(矢符で示すように)カテゴリの標準特徴ベクトルDが記憶されている。この標準特徴ベクトルDのサフィックスはカテゴリの番号であり、カテゴリがD1～DNまでN個あることを示している。標準特徴ベクトルDの要素であるdの()内の第1文字はカテゴリの番号を示し、第2文字は特徴要素の番号を示している。

【0025】次に、本実施の形態の動作を図8、図9、図10のフローチャートを用いて説明する。まず、特徴抽出部101は、カウンタKに「1」を設定し、変数Nに全カテゴリ(文字コード)数を設定する(S802)。次に、入力された学習用パターンから特徴ベクトル $F^K_i = (f^K_{i-1}, f^K_{i-2}, \dots, f^K_{i-n})$ を抽出する。ここで上付きのサフィックスは、カテゴリ番号を示す。下付きのサフィックスは、学習用パターンの同一カテゴリのi番目であることを示す。カテゴリ番号Kの学習用パターン数がM個あるときには特徴ベクトル $F^K_M$ まで抽出する(S804)。

【0026】次に標準特徴計算部102は、カテゴリ番号Kの標準特徴ベクトル $D^K = (d^K_1, d^K_2, \dots, d^K_n)$ を式(数1)を用いて計算する(S806)。次に、カウンタKの値が変数N以下か否かを判定し(S808)、肯定であれば、カウンタKに「1」を加えて(S810)、S804に戻り、否定であればS812に移る。

【0027】S812において、クラスタ代表辞書作成部103は、標準特徴ベクトル $D^1, D^2, \dots, D^N$ のクラスタ分析を行う。P個のクラスタ $C1_1 = \{D^{C1_1}_1, D^{C1_2}_1, \dots, D^{C1_{n1}}_1\}, \dots, C1_p = \{D^{Cp_1}_1, D^{Cp_2}_1, \dots, D^{Cp_{np}}_1\}$ に分類し(S814)、カウンタKに「1」を設定し(S816)、クラスタ $C1_K$ のクラスタ標準特徴 $C^K = (c^K_1, c^K_2, \dots, c^K_n)$ を式(数2)を用いて計算する(S902)。次に、カウンタKの値が全クラスタ数P以下か否かを判定し、肯定であればカウンタKに「1」を加えて(S906)、S902に戻り、否定であればクラスタ標準特徴 $C^1$ 、

$C^2, \dots, C^P$  を階層化認識辞書104に書き込む (S908)。

【0028】次に、クラスター一致度計算部401は、カウンタKに「1」を設定し (S910)、カウンタ1に「1」を設定する (S912)。カテゴリ標準特徴ベクトル $D^K$ とクラスター標準特徴ベクトル $C^1$ との一致度 $L_{11}$ を式(数3)を用いて計算する (S914)。カウンタ1の値がP以下であるか否かを判定し (S916)、肯定であればカウンタ1に「1」を加えて (S918)、S914に戻る。

【0029】否定であれば、クラスター内カテゴリ登録部402は、一致度Hの最大値、即ち、距離 $L$ の最小値となる $H_1 = \min (L_{11}, L_{12}, \dots, L_{1P})$ を求め (S920)、階層化認識辞書104のそのクラスター番号iのクラスター $C^1_i$ にカテゴリと標準特徴ベクトル $D^K$ と、併せて一致度Hを書き込む (S922)。カウンタKの値がカテゴリ数N以下であるか否かを判定し (S924)、肯定であればカウンタKに「1」を加えて (S926)、S912に戻る。

【0030】否定であれば、クラスター範囲設定部403は、カウンタKに「1」を設定し (S1002)、クラスター $C^1_K$ に分類されているカテゴリ $D^{CK,1}, \dots, D^{CK,mK}$ とそれぞれの一致度 $H^{K,1}, \dots, H^{K,mK}$ を読み出し (S1004)、一致度の最小値、即ち市街化距離の最大値をクラスター範囲 $r_K = \max (H^{K,1}, \dots, H^{K,mK})$ を求めクラスター範囲を設定する (S1006)。次にカウンタKの値がP以下か否かを判定し (S1008)、肯定であればカウンタKに「1」を加えて (S1010)、S1004に戻る。

【0031】否定であれば、クラスター拡張部404は、カウンタKに「1」を設定し (S1012)、クラスター標準特徴ベクトル $C^K$ を読み出し (S1014)、カウンタjに「1」を設定する (S1016)。次に、カテゴリの標準特徴ベクトル $D^j$ を読み出し (S1018)、両特徴ベクトル $C^K$ と $D^j$ との市街化距離 $L_1$ を式(数3)を用いて計算する (S1020)。次に、市街化距離 $L_1$ がクラスター設定範囲 $r_K$ 以下か否かを判定し (S1022)、肯定のときはクラスター $C^1_K$ にカテゴリを書き込み (S1024)、S1026に移る。これによって2つのクラスターの境界付近にあるカテゴリを2つのクラスターに重複して分類することができる。

【0032】否定のときは、S1026において、カウンタjの値がN以下か否かを判定し (S1026)、肯定のときはカウンタjに「1」を加えて (S1028)、S1018に戻る。否定のときはカウンタKの値がクラスター数P以下であるか否かを判定し (S1030)、否定のときはカウンタKに「1」を加えて (S1032)、S1014に戻る。

【0033】肯定のときは、クラスター内辞書作成部106は、階層化認識辞書104に記憶されているクラスター

に分類されているカテゴリの標準特徴ベクトルを全て書き込み (S1034)、処理を終了する。なお、本実施の形態において、クラスター範囲設定部403は各クラスターに対し、クラスターに登録されたカテゴリの中で最も一致度の低いカテゴリとの一致度をクラスター範囲として定めたが、クラスターの範囲は予め定めた他の数値でもよい。

【0034】以上のように本実施の形態によれば、パターン認識用辞書作成装置において、複数のカテゴリをグループ化し、クラスター標準特徴ベクトルを計算してクラスターの代表とし、各クラスター内に属するカテゴリを設定することにより高速なパターン認識装置を実現するための階層化辞書104を作成することができる。また、パターン認識用辞書作成装置のクラスター内カテゴリ登録部105において、クラスター内カテゴリの追加登録を行うことによりクラスター境界付近のカテゴリが複数のクラスターへ登録されることとなり、パターン認識装置におけるクラスター選択部によって選択されたクラスター内に正解となるカテゴリが含まれない事を防止し、認識率の低下を防ぐことができる。

(実施の形態2) 図11は、本発明に係るパターン認識用辞書作成装置の実施の形態2の構成図である。このパターン認識用辞書作成装置は、特徴抽出部101と、標準特徴計算部102と、クラスター代表辞書作成部103と、クラスター一致度計算部1101と、クラスター内カテゴリ登録部1102と、カテゴリ標準特徴ベクトル計算部1103と、階層化認識辞書1104とを備える。上記実施の形態1と同一の構成部分には実施の形態1と同一の符号を付してその説明を省略し、本実施の形態固有の構成部分についてのみ説明する。

【0035】クラスター一致度計算部1101は、クラスター代表辞書作成部103に起動されると、階層化認識辞書1104に記憶されているクラスター標準特徴ベクトル $C$ を読み出す。次に、特徴抽出部101に記憶されている学習用パターンの特徴ベクトル $F$ を1つ読み出し、クラスター標準特徴ベクトル $C$ と照合し、式(数6)によって一致度を順次計算する。

【0036】

【数6】

$$LC_i = \sum_{j=1}^n |f_j - c_{ij}| \quad (j=1, \dots, p)$$

ここでiはクラスター番号を示す。計算した一致度 $LC_i$ をクラスター内カテゴリ登録部1102にカテゴリ(文字コード)と特徴ベクトル $F$ とともに通知する。更に、クラスター内カテゴリ登録部1102から学習用パターンの特徴ベクトルの登録終了の通知を受けると、特徴抽出部101から学習用パターンの特徴ベクトル $F$ を更に1つ読み出し、一致度を計算し、計算した一致度をクラスター内カテゴリ登録部1102に上記と同様通知する。この



処理を入力された学習用パターンがなくなるまで繰り返す。

【0037】クラス内カテゴリ登録部1102は、クラスター一致度計算部1101から一致度LCIの通知を受けると、一致度LCIの最も高いクラスに特徴ベクトルを分類する。この分類に基づき、階層化認識辞書1104のクラスにその特徴ベクトルのカテゴリを書き込むとともに、そのカテゴリに対応付けて特徴ベクトルFを書き込む。書き込みが終了すると、クラスター一致度計算部1101に登録終了を通知する。全ての処理が終了するとカテゴリ標準特徴ベクトル計算部1103を起動する。

【0038】例えば、図3に示した文字コード「あ」の特徴ベクトルの一覧301が特徴抽出部101に記憶されている場合に、m1個の文字画像「あ」がクラスxに一致度が最も高く、m2個の文字画像「あ」がクラスyに一致度が最も高いとクラスター一致度計算部1101で計算され、クラス内カテゴリ登録部1102で分類されたときは、図12に示すように、階層化認識辞書1104に書き込まれる。

【0039】図12は、作成途中の階層化認識辞書1104を示し、階層化認識辞書1104は、クラスx欄1201とクラス標準特徴ベクトル欄1202とクラス内カテゴリ欄1203と特徴ベクトル欄1204、1205とを有する。特徴ベクトル欄1204には、クラスxに分類されたカテゴリ（文字コード「あ」）のm1個の特徴ベクトル $F_{x-1}, \dots, F_{x-m1}$ が記憶され、特徴ベクトル欄1205には、クラスyに分類されたカテゴリ（文字コード「あ」）のm2個の特徴ベクトル $F_{y-1}, \dots, F_{y-m2}$ が記憶されている。同一カテゴリ（文字コード）であっても、2つのクラスに分類されている。

【0040】カテゴリ標準特徴ベクトル計算部1103は、クラス内カテゴリ登録部1102に起動されると、階層化認識辞書1104に記憶されている特徴ベクトル欄1204、1205等を順次読み出し、同一欄（同一のクラスに分類されたカテゴリ）のカテゴリ標準特徴ベクトルを式(数1)を用いて計算する。なお、カテゴリ標準特徴ベクトル計算部1103で計算されたカテゴリ標準特徴ベクトルは、クラスに分類されたカテゴリの標準特徴ベクトルであり、標準特徴計算部1102で計算された標準特徴ベクトルが同一カテゴリ全ての標準特徴ベクトルであるのと異なる。

【0041】カテゴリ標準特徴ベクトル計算部1103は、階層化認識辞書1104の個々の特徴ベクトル欄1204、1205等のカテゴリ標準特徴ベクトルを計算すると、計算した特徴ベクトル欄を削除して、新たにカテゴリ標準特徴ベクトル欄1301（図13参照）を設けて、カテゴリ（文字コード）に対応付けてカテゴリ標準特徴ベクトルを書き込む。

【0042】階層化認識辞書1104は、RAM等からなり、図13に示すように、クラスx欄1201とクラス標準特徴ベクトル欄1202とクラス内カテゴリ欄1203とカテゴリ標準特徴ベクトル欄1301とを有する。クラスx欄1201には、各クラスのクラス番号が記憶されている。クラス標準特徴ベクトル欄1202には、クラス番号に対応したクラス標準特徴ベクトルが記憶されている。以上の欄1201、1202は、クラス代表辞書作成部103により書き込まれる。クラス内カテゴリ欄1203には、クラス標準特徴ベクトルに代表されて分類されるカテゴリ（文字コード）が記憶されている。この欄1203は、クラス内カテゴリ登録部1102により書き込まれる。

【0043】カテゴリ標準特徴ベクトル欄1301には、クラスに分類された学習パターンの特徴ベクトルの算術平均であるクラス内カテゴリ標準特徴ベクトルが記憶されている。この欄1301は、カテゴリ標準特徴ベクトル計算部1103により書き込まれる。上述（文字コード「あ」がクラスxとクラスyとの2つのクラスに重複して分類される。）したように、クラスxとクラスyとに文字コード「あ」が分類され、それらのカテゴリ標準特徴ベクトルは $D_k = (d(k, 1), \dots, d(k, n))$ 、 $D_s = (d(s, 1), \dots, d(s, n))$ とそれぞれ別の値として記憶されている。

【0044】次に、本実施の形態固有の動作を図14、図15のフローチャートを用いて説明する。まず、クラスター一致度計算部1101は、カウンタ1に「1」を設定し（S1402）、カテゴリ1の学習パターンの特徴ベクトル $F^1 = (f^1_{1-1}, \dots, f^1_{1-n})$ を全て読み出す（S1404）。カウンタmに「1」を設定し（S1406）、カウンタKに「1」を設定する（S1408）。特徴ベクトル $F_m$ とクラス標準特徴ベクトル $C_k$ との一致度（市街化距離） $LC_k$ を式(数6)を用いて計算する（S1410）。カウンタKの値がクラス数P以下か否かを判定する（S1412）。肯定であればカウンタKの値に「1」を加えて（S1413）、S1410に戻る。

【0045】否定であれば、クラス内カテゴリ登録部1102は、S1410で求めた市街化距離 $LC_k$ が最小値となるクラス番号qを求める（S1414）。クラスC1qにカテゴリ1を書き込み（S1416）、カテゴリ1の学習用パターンの特徴ベクトル $F_m$ をカテゴリ1に対応付けて書き込む（S1418）。次にカウンタmの値がカテゴリ1の学習用パターンの総数M1以下か否かを判定し（S1420）肯定であればカウンタmに「1」を加えて（S1422）、S1408に戻る。否定であれば、カウンタ1の値がカテゴリの総数N未満であれば、カウンタ1の値がカテゴリの総数N未満であるか否かを判定し（S1424）、肯定であればカ

ウンタ1に「1」を加えて(S1426)、1404に戻る。

【0046】否定であればカテゴリ標準特徴ベクトル計算部1103は、カウンタKに「1」を設定し1(S1502)、クラスタC<sub>1K</sub>に分類されているカテゴリを読み出し(S1504)、カウンタmに「1」を設定する(S1506)。次に、カテゴリD<sub>CLK,m</sub>に対応する特徴ベクトルF<sub>CLK,m</sub><sub>1</sub>, ..., F<sub>CLK,m</sub><sub>nK</sub>を読み出し(S1508)、クラスタK内のカテゴリ標準特徴ベクトルD<sub>CLK,m</sub>を式(数1)を用いて計算し、階層化認識辞書1104に書き込む(S1510)。カウンタmの値がクラスタC<sub>1K</sub>に分類されているカテゴリの総数n<sub>K</sub>以下か否かを判断し(S1512)、肯定のときはカウンタmに「1」を加えて(S1514)、S1508に戻り、否定のときはカウンタKの値がクラスタ総数P以下か否かを判定し(S1516)、肯定であればカウンタKに「1」を加えて(S1518)、S1504に戻り、否定であれば処理を終了する。以上のように本実施の形態によれば、同一カテゴリの学習用パターンが2個以上のクラスタに登録された場合には、各クラスタ毎にカテゴリ標準特徴ベクトルがそれぞれ計算され、カテゴリ標準特徴ベクトル欄1301に書き込まれる。そのため1つのカテゴリであってもクラスタが複数に亘ることとなり認識精度を向上させることができる。

【0047】なお、本実施の形態では、カテゴリ特徴ベクトル計算部1103において、各クラスタに分類された同一カテゴリの特徴ベクトルに基づいて、各クラスタごとのカテゴリの標準特徴ベクトルを求めたけれども、本発明の他の実施の形態として、標準特徴計算部102で計算された標準特徴ベクトルを複数のクラスタの同一カテゴリの標準特徴ベクトルとしてもよい。

【0048】(実施の形態3) 図16は、本発明に係るパターン認識装置の実施の形態3の構成図である。このパターン認識装置は、階層化認識辞書104と、特徴抽出部1601と、クラスター一致度計算部1602と、クラスタ選択部1603と、クラスタ内カテゴリ照合部1603と、クラスタ内カテゴリ照合部1604と、カテゴリ決定部1605とを備えている。

【0049】階層化認識辞書104は、ROM等からなり、実施の形態1のパターン認識用辞書作成装置で作成されたものである。図7に示すように、クラスタ705とカテゴリ706とが階層的に構成され、クラスタを代表するクラスタ標準特徴ベクトル707とカテゴリを代表する標準特徴ベクトル708とがそれぞれクラスタとカテゴリとに対応付けられている。

【0050】特徴抽出部1601は、入力された未知のパターンから特徴ベクトルF=(f<sub>1</sub>, f<sub>2</sub>, ..., f<sub>n</sub>)を抽出して、クラスター一致度計算部1604に通知する。クラスター一致度計算部1602は、階層化認識辞書104からクラスタ欄702に示されるP個のクラスタのク

ラスタ標準特徴ベクトル707との一致度を市街化距離LC<sub>i</sub>の式(数6)を用いて計算する。P個のクラスタについて一致度を求めると、クラスタ選択部1603を起動する。

【0051】クラスタ選択部1603は、クラスター一致度計算部1603に起動されると、クラスター一致度計算部1602で求められた各クラスタとの市街化距離LC<sub>i</sub>をソートし、値の小さい方から並べ、例えば予め定めたs個のクラスタを候補クラスタとして選択する。このとき、クラスタ総数の5~10%のクラスタを候補クラスタとして選択してもよい。候補クラスタの選択が終了するとクラスタ内カテゴリ照合部1604を起動する。

【0052】クラスタ内カテゴリ照合部1604は、クラスタ選択部1603に起動されると、候補クラスタとされたs個のクラスタに分類された各カテゴリの標準特徴ベクトルDを順次読み出して、特徴ベクトルFと市街化距離LDを式(数7)を用いて計算する。

【0053】

【数7】

$$LD = \sum_{i=1}^n |f_i - d_i|$$

候補クラスタに選択された全てのカテゴリとの一致度の計算が終了するとカテゴリ決定部1605を起動する。カテゴリ決定部1605は、クラスタ内カテゴリ照合部1604に起動されるとクラスタ内カテゴリ照合部1604で得られた市街化距離LDの最小となる(一致度の最も高い)標準特徴ベクトルに対応するカテゴリを階層化認識辞書104から読み出し、認識結果と決定して出力する。

【0054】例えば、クラスタ選択部1603で、予め定められたSが「2」の場合に、クラスタiとクラスタjとが候補クラスタとして選択されたときには(図7参照)、クラスタ内カテゴリ欄701に登録されているクラスタiの文字「責」、「草」、…、「軍」の標準特徴ベクトル「D<sub>h</sub>」…「D<sub>k</sub>」及びクラスタjの文字コード「軍」、「間」、…、「開」の標準特徴ベクトルとの一致度が計算される。

【0055】次に、本実施の形態の動作を図17、図18のフローチャートを用いて説明する。まず、特徴抽出部1601は、入力パターンを受け取ると、特徴ベクトルF=(f<sub>1</sub>, ..., f<sub>n</sub>)を抽出する(S1702)。クラスター一致度計算部1602は、階層化認識辞書104からクラスタ数をP読み出し、カウンタKに「1」を設定する(S1704)。次にクラスタKのクラスタ標準特徴ベクトルを読み出し、特徴ベクトルFとの市街化距離LC<sub>K</sub>を式(数6)を用いて計算する(S1706)。カウンタKの値がクラスタ総数P以下か否かを判定し(S1708)、肯定のときはS1706に戻る。

【0056】否定であるときは、S1706で求めた市街化距離LC<sub>K</sub>をソートし(1712)、距離の小さい

方から並べる (S1716)。値の小さい方からクラスタを  $s$  個選択する (S1716)。次に、クラスタ内カテゴリ照合部1604は、カウンタ  $K$  に「1」を設定し (1806)、カウンタ  $t$  に「0」を設定する (S1802)。選択されたクラスタ  $C_k$  内のカテゴリの標準特徴ベクトル  $D_{k1}, \dots, D_{k, m_k}$  を読み出し ( $m_k$  はクラスタ  $C_k$  内のカテゴリの総数である。) (S1804)、カウンタ  $j$  に「1」を設定し (S1806)、カテゴリ  $D_{kj}$  との市街化距離  $LD(t+j)$  を式 (数7) を用いて計算する。次にカウンタ  $j$  の値がカテゴリの総数  $m_k$  以下であるか否かを判定し (S1810)、肯定のときはカウンタ  $j$  に1を加えて (S1812)、S1808に戻る。否定のときはカウンタ  $K$  の値が選択数  $S$  以下であるか否かを判定し (S1814)、肯定のときはカウンタ  $t$  に「 $mK$ 」を加え、(S1816)、カウンタ  $K$  に「1」を加えて (S1818)、S1804に戻る。

【0057】否定のときは、カテゴリ決定部1605は、S1808で得られた市街化距離  $LD1, LD2, \dots, LDt$  の最小値に対応するカテゴリを選択する (S1820)。選択されたカテゴリを認識結果として出力する (S1822)。なお、本実施の形態のクラスタ選択部1603において一致度の高いものから一定個数のクラスタを選択するとしたが、一致度が予め定めた一定値以上 (距離値  $LC$  が一定値以下) のクラスタを選択することもできる。

【0058】また、本実施の形態のクラスタ選択部1603においては、一致度の高いものから順次クラスタを選択し、選択されたクラスタ内カテゴリ数を順次計算し、カテゴリ数が予め定めた一定個数例えば、全カテゴリ数の20%に到達するまでクラスタ数を選択してクラスタ選択をすることもできる。これによって、カテゴリの標準特徴との照合数を規制して処理時間を適正とすることができる。

【0059】更に、本実施の形態のカテゴリ決定部1605は、クラスタ内カテゴリ照合部1604によって求められた各カテゴリに対する一致度の中で最も一致度の高いカテゴリを認識結果としたけれども、他の実施の形態として、クラスタ一致度計算部1602で求められたクラスタ標準特徴ベクトルとの一致度  $LC$  とクラスタ内カテゴリ照合部1604で求められたカテゴリの標準特徴ベクトルとの一致度  $LD$  との両方を統合して、以下のように認識結果を決定することもできる。

【0060】例えば、入力パターンの特徴ベクトルを  $F$ 、クラスタ選択部1603において選択されたクラスタ  $i$  との距離値を  $LC$ 、クラスタ  $i$  内のカテゴリである文字コード「 $g$ 」に対する距離値を  $LDh$  とすると、クラスタ  $i$  内のカテゴリである文字コード「 $g$ 」に対する一致度  $W$  を

$$W = LC + a \times LD \quad (a \text{ は予め定めた定数})$$

のように定義することにより、クラスタとの一致度と各カテゴリとの一致度を統合した総合一致度を計算し、最も総合一致度の高いカテゴリを認識結果とする。

【0061】以上のように本実施の形態によれば、入力パターンの特徴ベクトルと階層化認識辞書104に記憶されているクラスタ標準特徴ベクトルとの照合を行い、クラスタ選択部1603によって選択されたクラスタ内のカテゴリの標準特徴ベクトルのみとの照合処理を行うことによって、不要なカテゴリとの照合処理を減少させることができるため高速なパターン認識装置を実現できる。

【0062】なお、本実施の形態では、図16に示したような構成で本発明に係るパターン認識装置を実現したけれども、本発明はプログラムによって実現し、これをフロッピーディスク等の記録媒体に記録して移送することにより、独立した他のコンピュータ・システムで容易に実施することができる。図19は、これをフロッピーディスクで実施する場合を説明する図である。

【0063】記録媒体本体であるフロッピーディスク1901の物理フォーマットは、同心円状に外周から内周に向かってトラック1、2、 $\dots$ 、80を作成し、角度方向に16のセクタに分割している。このように割り当てられた領域に従って、プログラムを記録する。このフロッピーディスク1901は、ケース1902に収納され、これによって、ディスクを埃や外部からの衝撃から守り、安全に移送することができる。

【0064】図20は、フロッピーディスク1901にプログラムの記録再生を行うことを説明する図である。図示のようにコンピュータ・システム2003にフロッピーディスクドライブ2001を接続することにより、ディスク1901に対してプログラムを記録再生することが可能となる。ディスク1901はフロッピーディスクドライブ2001に、挿入口2002を介して組込み、および取り出しがなされる。記録する場合はコンピュータ・システム2003からプログラムをフロッピーディスクドライブ2001によってディスク1901に記録する。再生する場合は、フロッピーディスクドライブ2001がプログラムをディスク1901から読み出し、コンピュータ・システム2003に転送する。

【0065】なお、この実施の形態においては、記録媒体としてフロッピーディスクを用いて説明を行ったが、光ディスクを用いても同様に行うことができる。また記録媒体はこれらに限られず、ICカード、ROMカセット等、プログラムを記録できるものであれば、同様に実施することができる。

【0066】

【発明の効果】以上のように、本発明によれば、カテゴリごとに入力される複数の学習用パターンからそのカテゴリの認識に用いる特徴量を抽出する特徴量抽出手段と、前記特徴量抽出手段で抽出された特徴量に基づいて

各カテゴリごとの標準特微量を計算する標準特微量計算手段と、前記標準特微量計算手段で計算された標準特微量をクラスタ分析して、複数のクラスタに各カテゴリを分類するカテゴリ分類手段とを有するパターン認識用辞書作成装置であって、前記カテゴリ分類手段で分類された各カテゴリの標準特微量に基づいてクラスタのクラスタ標準特微量を計算するクラスタ標準特微量計算手段と、前記クラスタ標準特微量計算手段で計算されたクラスタのクラスタ標準特微量と、前記カテゴリ分類手段で該クラスタに分類されたカテゴリと、前記標準特微量計算手段で計算されたそのカテゴリの標準特微量とを対応して階層化認識辞書に登録する登録手段とを備えるよう構成したので、各カテゴリの標準特微量を分類した複数のクラスタの各クラスタの標準特微量をカテゴリ決定のための第1次の被認識対象とすることができる階層化認識辞書が得られる。これによって、照合対象数を減らすことができ、認識結果たるカテゴリを短時間で決定できるパターン認識装置が実現できる。

【0067】また、本発明によれば、カテゴリごとに入力される複数の学習用パターンからそのカテゴリの認識に用いる特微量を抽出する特微量抽出手段と、前記特微量抽出手段で抽出された特微量に基づいて各カテゴリごとの標準特微量を計算する標準特微量計算手段と、前記標準特微量計算手段で計算された標準特微量をクラスタ分析して、複数のクラスタに分類する分類手段とを有するパターン認識用辞書作成装置であって、前記分類手段で分類された標準特微量に基づいて、各クラスタごとのクラスタ標準特微量を計算するクラスタ標準特微量計算手段と、前記標準特微量計算手段で計算されたカテゴリの標準特微量と前記クラスタ標準特微量計算手段で計算された各クラスタのクラスタ標準特微量との一致度を計算する一致度計算手段と、前記一致度計算手段で計算された一致度の値の最も高いクラスタにカテゴリを分類するカテゴリ分類手段と、前記カテゴリ分類手段で分類されたカテゴリと該カテゴリの標準特微量とを前記クラスタ標準特微量とに対応して階層化認識辞書に登録する登録手段とを備えるよう構成したので、各クラスタに一致度の高いカテゴリを分類するので、カテゴリ決定のための第1次の被認識対象たるクラスタ標準特微量からクラスタ候補を選択すると、第2次の被認識対象たるカテゴリを標準特微量から精度よく決定することが可能な階層化認識辞書を作成できる。

【0068】また、本発明によれば、前記階層化認識辞書に登録されているクラスタ内のカテゴリのうち一致度の値の最も低い値を該クラスタのクラスタ範囲と設定するクラスタ範囲設定手段と、前記クラスタ範囲設定手段で設定されたクラスタのクラスタ範囲内に前記カテゴリ分類手段で他のクラスタに分類されたカテゴリがあるときには、階層化認識辞書の該クラスタに重複して該カテゴリとその標準特微量とを対応付けて追加登録する追加

登録手段とを備えるよう構成したので、クラスタに重複してカテゴリを分類するので、クラスタ境界領域付近にあるカテゴリを他のクラスタ候補からも認識結果とする階層化認識辞書を得ることができる。記載のパターン認識用辞書作成装置。

【0069】また、本発明によれば、前記追加登録手段は、前記クラスタ標準特微量計算手段で計算されたクラスタの標準特微量と、前記標準特微量計算手段で計算された各標準特微量との一致度を計算する一致度計算部と、前記一致度計算部で計算された一致度が当該クラスタのクラスタ範囲内に前記カテゴリ分類手段で他のクラスタに分類された標準特微量を有するカテゴリがあるかを判定するクラスタ範囲判定部と、前記クラスタ範囲判定部があると判定したとき、当該カテゴリとその標準特微量とを対応付けて当該クラスタに登録する重複登録部とを有するよう構成しているので、カテゴリを確実に複数クラスタに重複登録した階層化認識辞書を得ることができる。

【0070】また、本発明によれば、カテゴリごとに入力される複数の学習用パターンからそのカテゴリの認識に用いる特微量を抽出する特微量抽出手段と、前記特微量抽出手段で抽出された特微量に基づいて各カテゴリごとの標準特微量を計算する標準特微量計算手段と、前記標準特微量計算手段で計算された標準特微量をクラスタ分析して、複数のクラスタに分類する分類手段とを有するパターン認識用辞書作成装置であって、前記分類手段で分類された標準特微量に基づいて、各クラスタごとのクラスタ標準特微量を計算するクラスタ標準特微量計算手段と、前記特微量抽出手段で抽出された特微量と前記クラスタ標準特微量計算手段で計算された各クラスタのクラスタ標準特微量との一致度を計算する学習パターンクラスタ一致度計算手段と、前記学習パターンクラスタ一致度計算手段で計算された一致度の最も高いクラスタに当該カテゴリと特微量とを分類する特微量分類手段と、前記特微量分類手段で分類された同一カテゴリの特微量に基づいて各カテゴリごとの当該クラスタにおけるクラスタ内標準特微量を計算するクラスタ内標準特微量計算手段と、前記カテゴリ分類手段で分類されたカテゴリと前記クラスタ内標準特微量計算手段で計算された該カテゴリのクラスタ内標準特微量とを前記クラスタ標準特微量とに対応付けて階層化認識辞書に登録する登録手段とを備えるよう構成しているので、分類されたクラスタにカテゴリを分類する際に、個々の学習用パターンから抽出された特微量との一致度を用いるので、学習用パターンに拡がり（多様なパターン）がある場合にも、十分その特微量を代表するクラスタにカテゴリを分類し、またその分類したクラスタのカテゴリごとの特微量に基づいて第2次の被認識対象たる標準特微量を精度よく得ることができる。

【0071】また、本発明によれば、カテゴリごとに入

力される複数の学習用パターンからそのカテゴリの認識に用いる特徴量を抽出する特徴量抽出手段と、前記特徴量抽出手段で抽出された特徴量に基づいて各カテゴリごとの標準特徴量を計算する標準特徴量計算手段と、前記標準特徴量計算手段で計算された標準特徴量をクラスタ分析して、複数のクラスタに分類する分類手段とを有するパターン認識用辞書作成装置であって、前記分類手段で分類された標準特徴量に基づいて、各クラスタごとのクラスタ標準特徴量を計算するクラスタ標準特徴量計算手段と、前記特徴量抽出手段で抽出された特徴量と前記クラスタ標準特徴量計算手段で計算された各クラスタのクラスタ標準特徴量との一致度を計算する学習パターン-クラスタ一致度計算手段と、前記学習パターン-クラスタ一致度計算手段で計算された一致度の最も高いクラスタに当該カテゴリを分類する特徴量分類手段と、前記特徴量分類手段で分類されたカテゴリと、該カテゴリの前記標準特徴量計算手段で計算された標準特徴量とを前記クラスタ標準特徴量に対応付けて階層化認識辞書に登録する登録手段とを備えるよう構成しているため、分類されたクラスタにカテゴリを分類する際に、個々の学習用パターンから抽出された特徴量との一致度を用いるので、学習用パターンにばらつきがある場合にも、十分その特徴量を代表するクラスタにカテゴリを分類することができる。

【0072】また、本発明によれば、前記クラスタ標準特徴量計算手段は、同一クラスタ内の全カテゴリの標準特徴量の各要素を算術平均してクラスタ標準特徴量の各要素とするようしているため、クラスタを代表する特徴量を正しく把握することができる。また、本発明によれば、上記一致度の計算は、市街化距離、二乗距離又は類似度のいずれかにより計算されるようしているため、一致度を正確に把握することができる。

【0073】また、本発明によれば、入力パターンをカテゴリとして認識するパターン認識装置であって、各カテゴリの学習用パターンの特徴量に基づいて計算された標準特徴量とカテゴリとを、当該標準特徴量のクラスタ分析によって複数のクラスタに分類して、分類したクラスタと、クラスタに属するカテゴリの標準特徴量に基づいて計算されたクラスタ標準特徴量とに対応付けて登録している階層化認識辞書と、入力パターンから認識に用いる特徴量を抽出する特徴量抽出手段と、前記特徴量抽出手段で抽出された特徴量と前記階層化認識辞書に登録されているクラスタ標準特徴量との一致度を計算するクラスタ一致度計算手段と、前記クラスタ一致度計算手段で計算された一致度に基づいて候補クラスタを選択する候補クラスタ選択手段と、前記候補クラスタ選択手段で選択された候補クラスタに対応付けられた標準特徴量と前記特徴量抽出手段で抽出された特徴量との一致度を計算するカテゴリ一致度計算手段と、前記カテゴリ一致度計算手段で計算された一致度に基づいて認識結果である

カテゴリを決定するカテゴリ決定手段とを備えるよう構成しているため、第1の被認識対象としてクラスタ標準特徴量と入力パターンの特徴量とから計算した一致度に基づいて候補クラスタを選択して、選択されたクラスタ内の第2の被認識対象として標準特徴量と入力パターンの特徴量とから計算した一致度に基づいて認識結果たるカテゴリを決定できるので、照合に要する時間を短縮した効率的なパターン認識装置を得ることができる。

【0074】また、本発明によれば、前記候補クラスタ選択手段は、前記クラスタ一致度計算手段で計算された一致度に基づいて、一致度の高いクラスタから順に所定個数のクラスタを候補クラスタとして選択する所定数候補クラスタ選択部を有するよう構成しているため、所定個数の候補クラスタに分類されたクラスタ内のカテゴリの標準特徴量と入力パターンの特徴量との一致度だけを計算することで、認識結果たるカテゴリを精度よく決定できるので、認識処理に要する処理時間を短縮することができる。

【0075】また、本発明によれば、前記候補クラスタ選択手段は、前記クラスタ一致度計算手段で計算された一致度に基づいて、一致度が予め定めたいきい値以上のクラスタを候補クラスタとして選択するしきい値候補クラスタ選択部を有するよう構成しているため、少ないカテゴリの標準特徴量との一致度の計算により、精度よくカテゴリを決定することができる。

【0076】また、本発明によれば、前記候補クラスタ選択手段は、前記クラスタ一致度計算手段で計算された一致度に基づいて、一致度の最も高いクラスタの一致度を基準値として基準値との差又は比が一定値以内の一致度を有するクラスタを候補クラスタとして選択する基準値候補クラスタ選択部を有するよう構成しているため、候補クラスタの選択漏れが防止され、精度のよいカテゴリの決定が短時間でできる。

【0077】また、本発明によれば、前記候補クラスタ選択手段は、前記クラスタ一致度計算手段で計算された一致度の高いクラスタから順にクラスタ内に登録されているカテゴリ数が予め定めたカテゴリ数に到達するまでのクラスタを候補クラスタとして選択するカテゴリ数候補クラスタ選択部を有するよう構成しているため、入力パターンの特徴量と一致度を計算するカテゴリの標準特徴量の数を規制することによって、処理時間を短縮して認識の効率化を図ることができる。

【0078】また、本発明によれば、前記カテゴリ決定手段は、前記クラスタ選択手段で選択された候補クラスタの前記クラスタ一致度計算手段で計算された一致度と、前記カテゴリ一致度計算手段で計算された一致度とに基づいて総合一致度を計算する総合一致度計算部と、前記総合一致度計算部で計算された総合一致度に基づいてカテゴリを決定する総合カテゴリ決定部とを有するよう構成しているため、クラスタの一致度とカテゴリの一

致度との両方を加味してカテゴリを決定し、更に精度よく短時間にカテゴリを決定できる。

【0079】更に、本発明によれば、階層化認識辞書を用いて入力パターンをカテゴリとして認識するパターン認識方法が記録された記録媒体であって、各カテゴリの学習用パターンの特徴量に基づいて計算された標準特徴量とカテゴリとを、当該標準特徴量のクラス分析によって複数のクラスに分類して、分類したクラスと、クラスに属するカテゴリの標準特徴量に基づいて計算されたクラス標準特徴量とに対応付けて登録している階層化認識辞書と、以下の特定ステップを実行するプログラムとが記録され、特定ステップには、入力パターンから認識に用いる特徴量を抽出する特徴量抽出ステップと、前記特徴量抽出ステップで抽出された特徴量と前記階層化認識辞書に登録されているクラス標準特徴量との一致度を計算するクラス一致度計算ステップと、前記クラス一致度計算ステップで計算された一致度に基づいて候補クラスを選択する候補クラス選択ステップと、前記候補クラス選択ステップで選択された候補クラスに対応付けられた標準特徴量と前記特徴量抽出ステップで抽出された特徴量との一致度を計算するカテゴリ一致度計算ステップと、前記カテゴリ一致度計算ステップで計算された一致度に基づいて認識結果であるカテゴリを決定するカテゴリ決定ステップとを含むこととしているので、パターン認識機能を有しない他のパーソナルコンピュータと相俟って、上記パターン認識装置と同様の効果を容易に得ることができる。

#### 【図面の簡単な説明】

【図1】本発明に係るパターン認識用辞書作成装置の実施の形態1の構成図である。

【図2】上記実施の形態において入力される学習用パターンの一例を示す説明図である。

【図3】上記実施の形態の特徴抽出部で抽出された特徴ベクトルの一例を示す説明図である。

【図4】上記実施の形態のクラス内のカテゴリ設定部の詳細な構成図である。

【図5】上記実施の形態のクラス内カテゴリ登録部の処理内容の説明図である。

【図6】上記実施の形態のクラス範囲設定部でのクラス範囲の設定の説明図である。

【図7】上記実施の形態の階層化認識辞書の一例を示す説明図である。

【図8】上記実施の形態の動作を説明するフローチャートである。

【図9】上記実施の形態の動作を説明するフローチャートである。

【図10】上記実施の形態の動作を説明するフローチャートである。

【図11】本発明に係るパターン認識用辞書作成装置の実施の形態2の構成図である。

【図12】上記実施の形態の階層化認識辞書の作成途中の状態を示す説明図である。

【図13】上記実施の形態の階層化認識辞書の記憶内容の説明図である。

【図14】上記実施の形態の動作を説明するフローチャートである。

【図15】上記実施の形態の動作を説明するフローチャートである。

【図16】本発明に係るパターン認識装置の実施の形態3の構成図である。

【図17】上記実施の形態の動作を説明するフローチャートである。

【図18】上記実施の形態の動作を説明するフローチャートである。

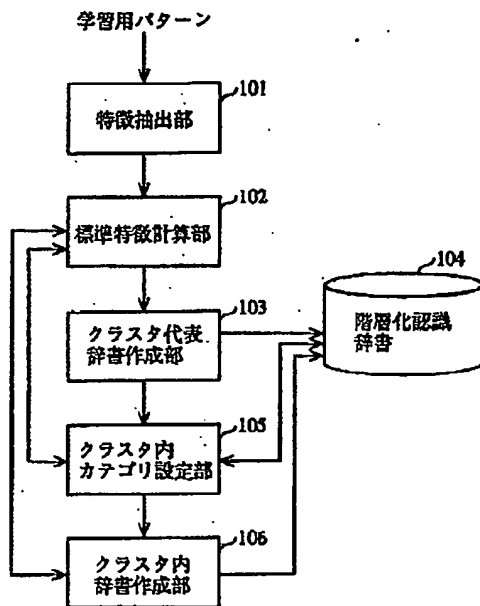
【図19】本発明に係るパターン認識方法のプログラムを記録した記録媒体の構成を示す説明図である。

【図20】上記記録媒体のコンピュータシステムへの装着を説明する図である。

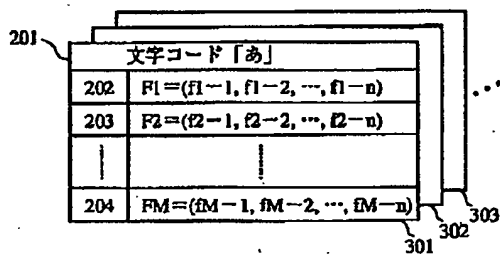
#### 【符号の説明】

|             |                |
|-------------|----------------|
| 101         | 特徴抽出部          |
| 102         | 標準特徴計算部        |
| 103         | クラス代表辞書作成部     |
| 104、1104    | 階層化認識辞書        |
| 105         | クラス内カテゴリ設定部    |
| 106         | クラス内辞書作成部      |
| 30 401、1101 | クラス一致度計算部      |
| 402、1102    | クラス内カテゴリ登録部    |
| 403         | クラス範囲設定部       |
| 404         | クラス拡張部         |
| 702、1201    | クラス欄           |
| 703、1202    | クラス標準特徴ベクトル欄   |
| 1103        | カテゴリ標準特徴計算部    |
| 701、1203    | クラス内カテゴリ欄      |
| 704         | 標準特徴ベクトル欄      |
| 1204、1205   | 特徴ベクトル欄        |
| 40 1301     | カテゴリ標準特徴ベクトル欄  |
| 1601        | 特徴抽出部          |
| 1602        | クラス一致度計算部      |
| 1603        | クラス選択部         |
| 1604        | クラス内カテゴリ一致度計算部 |
| 1605        | カテゴリ決定部        |

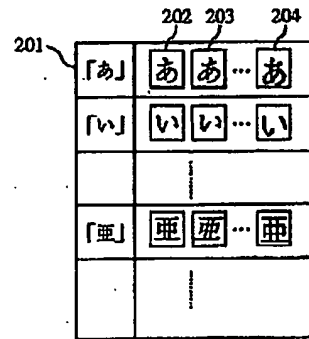
【図1】



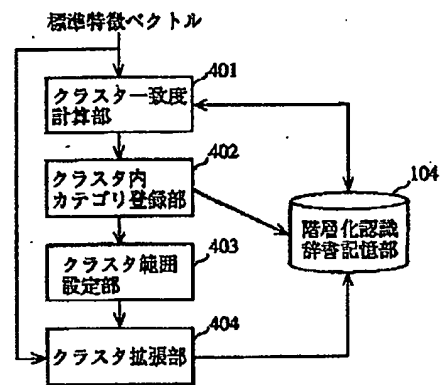
【図3】



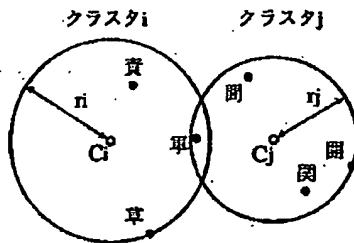
【図2】



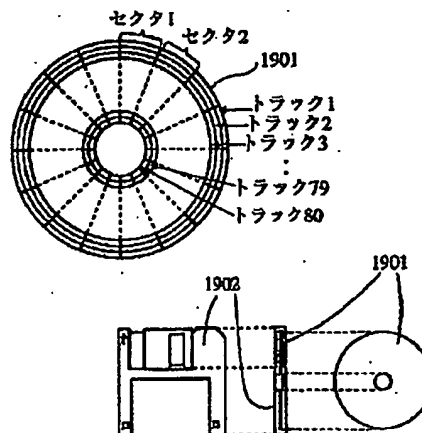
【図4】



【図6】



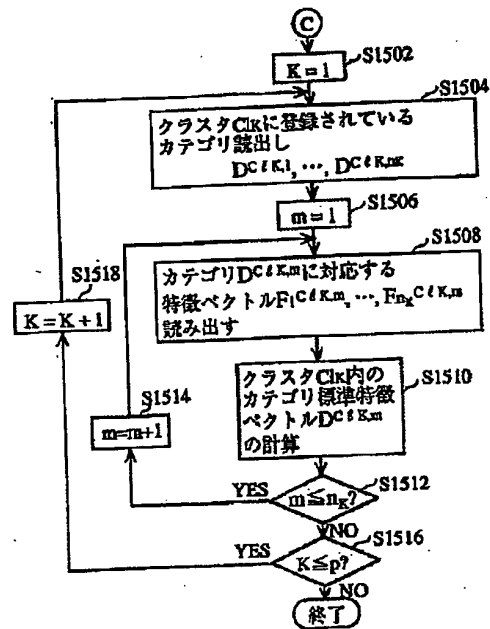
【図19】



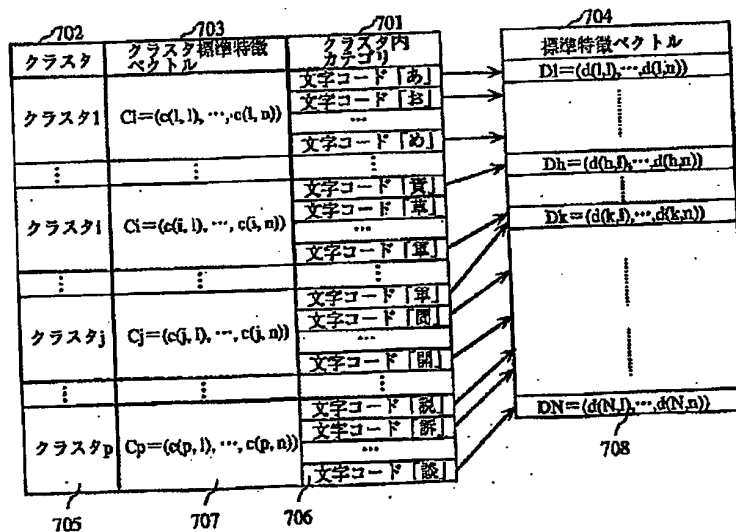
【図5】

| 501<br>クラス | 502<br>クラス内カテゴリ | 503<br>一致度 |
|------------|-----------------|------------|
| クラス1       | カテゴリ「あ」         | H1-1       |
|            | カテゴリ「お」         | H1-2       |
|            | ...             |            |
| クラス2       | カテゴリ「め」         | H1-m1      |
|            | カテゴリ「開」         | H2-1       |
|            | カテゴリ「関」         | H2-2       |
| ...        | ...             |            |
|            | カテゴリ「関」         | H2-m2      |
| クラスp       | カテゴリ「説」         | Hp-1       |
|            | カテゴリ「訴」         | Hp-2       |
|            | ...             |            |
| ...        | カテゴリ「談」         | Hp-m       |

【図16】

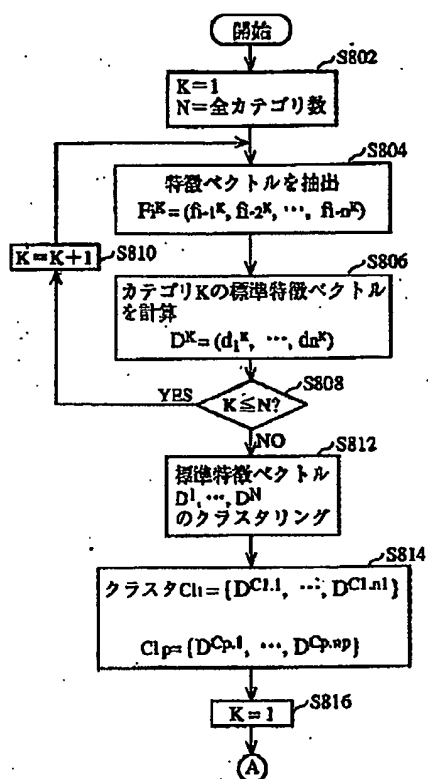


【図7】

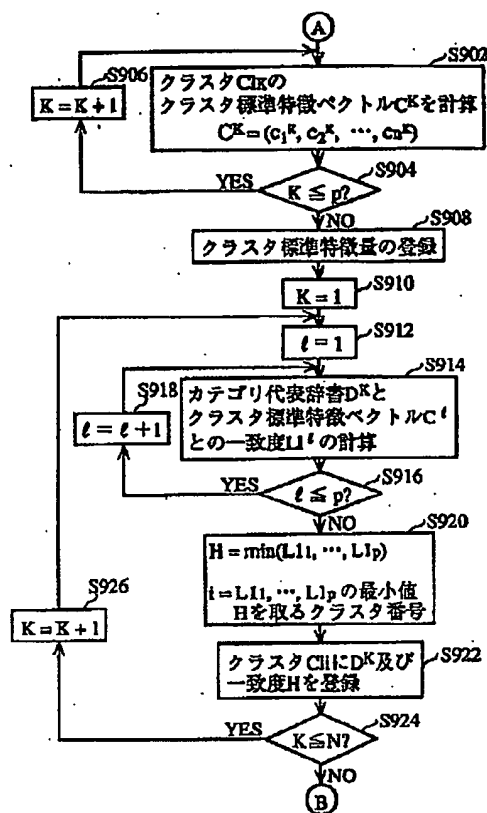




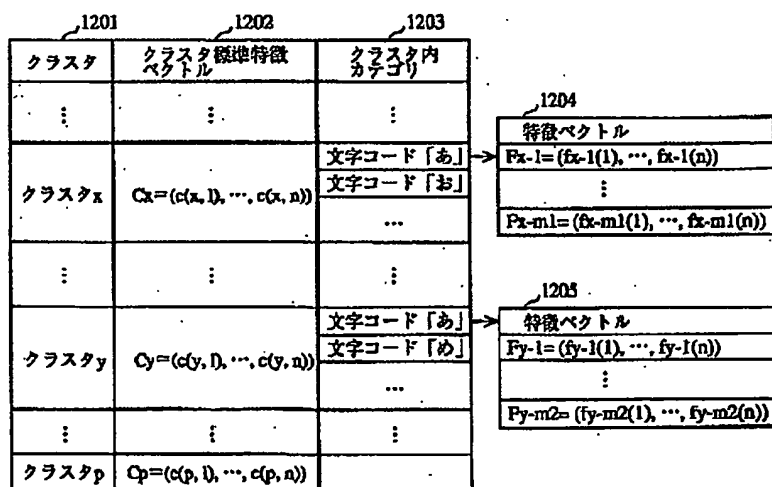
【図8】



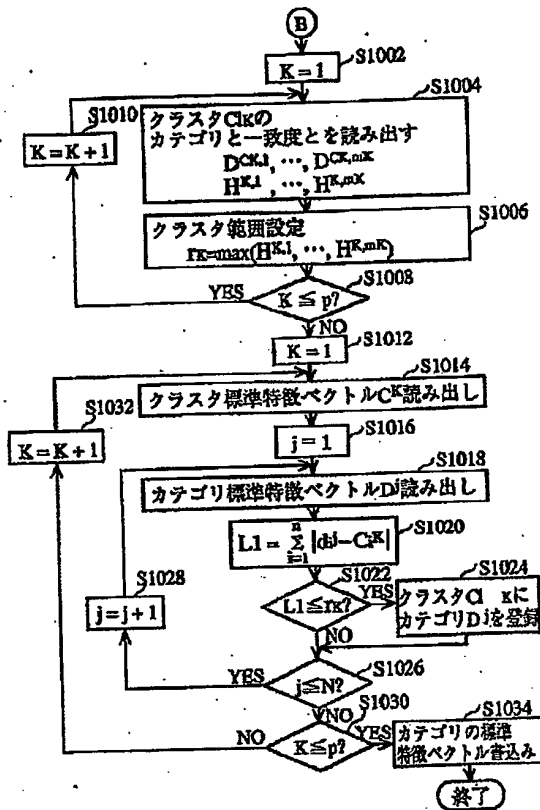
【図9】



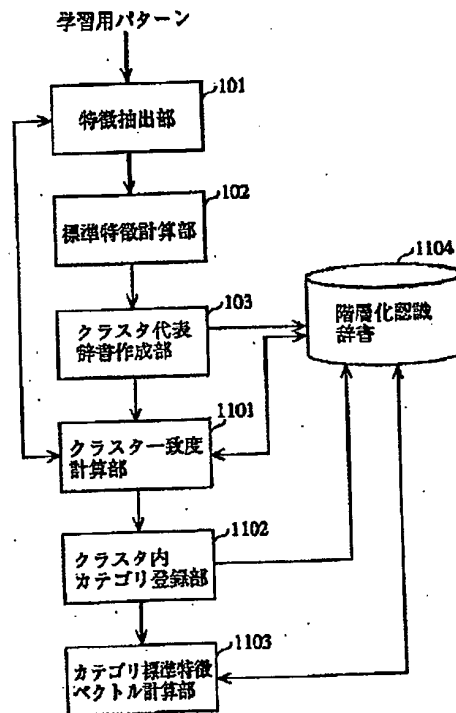
【図12】



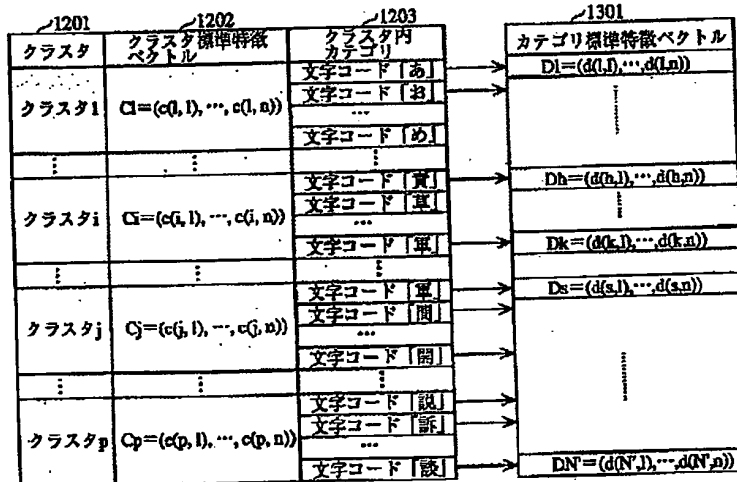
【図10】



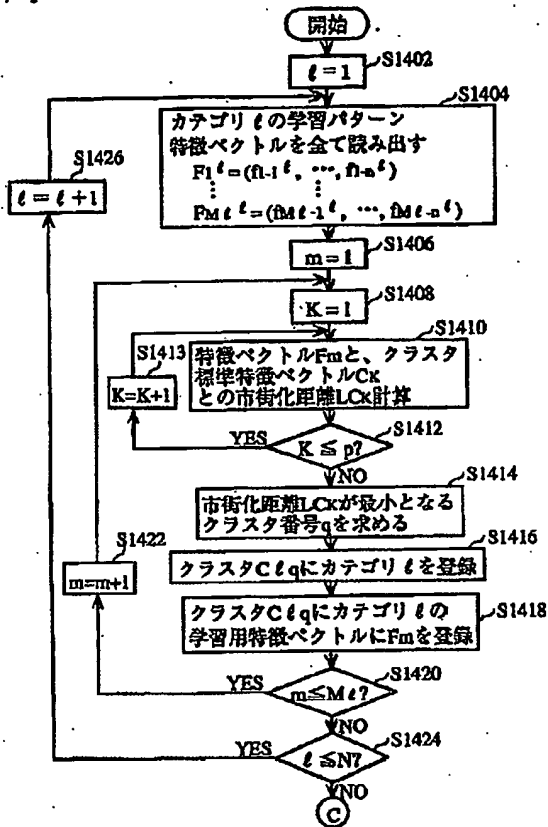
【図11】



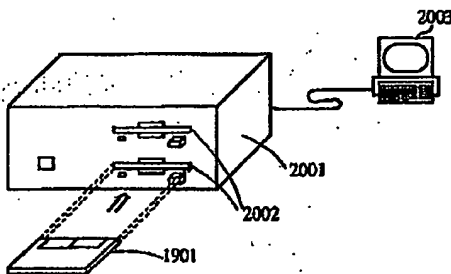
【図13】



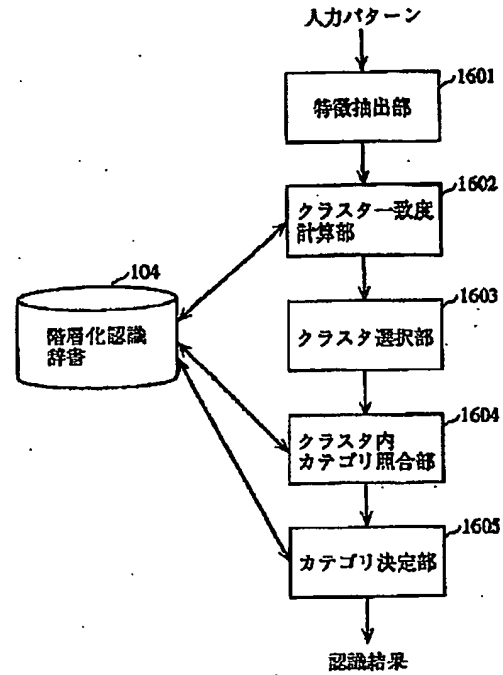
【図14】



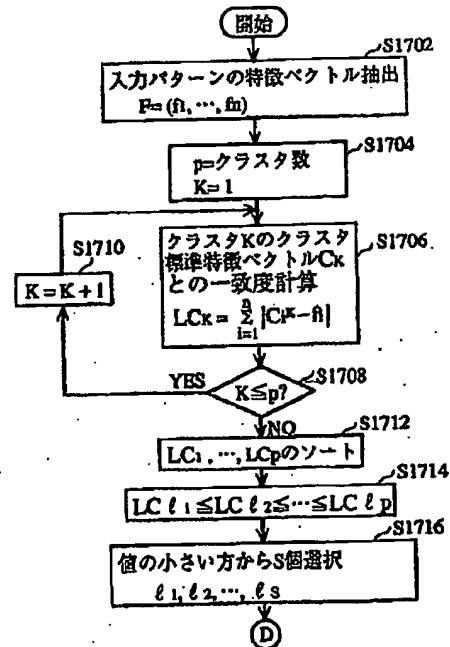
【図20】



【図16】



【図17】



【図18】

